An Open-Source Hierarchical Phrase-Based Machine Translation System

Lane Schwartz
lane@cs.umn.edu

University of Minnesota
Background

Statistical Methods

Hierarchical phrase-based translation
Translation is one of the oldest problems in computer science.
Background

Translation is one of the oldest problems in computer science.

- Rule-based translation
Translation is one of the oldest problems in computer science.

- Rule-based translation

  Requires highly trained bilingual linguists
Background

Translation is one of the oldest problems in computer science.

- Rule-based translation

  Requires highly trained bilingual linguists
  And lots and lots of time for them to develop rules
Background

Translation is one of the oldest problems in computer science.

- Rule-based translation

  Requires highly trained bilingual linguists
  And lots and lots of time for them to develop rules
  Hard to extend to new domains
Translation is one of the oldest problems in computer science.

- Rule-based translation
  Requires highly trained bilingual linguists
  And lots and lots of time for them to develop rules
  Hard to extend to new domains

- Statistical translation
Translation is one of the oldest problems in computer science.

- Rule-based translation
  
  Requires highly trained bilingual linguists
  And lots and lots of time for them to develop rules
  Hard to extend to new domains

- Statistical translation ← We’ll use this approach :)}
Parallel Corpora

Statistical translation requires large parallel texts.
Parallel Corpora

Statistical translation requires large parallel texts.

- Corporate documentation
Parallel Corpora

Statistical translation requires large parallel texts.

- Corporate documentation
- International news organizations
Parallel Corpora

Statistical translation requires large parallel texts.

- Corporate documentation
- International news organizations
- Governments
Parallel Corpora

Statistical translation requires large parallel texts.

- Corporate documentation
- International news organizations
- Governments
  - United Nations
Parallel Corpora

Statistical translation requires large parallel texts.

- Corporate documentation
- International news organizations
- Governments
  - United Nations
  - Canada
Parallel Corpora

Statistical translation requires large parallel texts.

- Corporate documentation
- International news organizations
- Governments
  - United Nations
  - Canada
  - European Union
Parallel Corpora

Statistical translation requires large parallel texts.

- Corporate documentation
- International news organizations
- Governments
  - United Nations
  - Canada
  - European Union ← We’ll use this data :)
Words

- Word-by-word translation
Words

- Word-by-word translation
- Word alignments
Words

- Word-by-word translation
- Word alignments
  - Align words by hand
Words

- Word-by-word translation
- Word alignments
  - Align words by hand
  - Tedious and time-consuming
Words

- Word-by-word translation
- Word alignments
  - Align words by hand
tedious and time-consuming
  - Automatic alignment
Words

- Word-by-word translation
- Word alignments
  - Align words by hand
tedious and time-consuming
  - Automatic alignment
uses EM on a parallel corpus;
see Och & Ney (2000)
Words

- Word-by-word translation
- Word alignments
  - Align words by hand
tedious and time-consuming
  - Automatic alignment
    uses EM on a parallel corpus;
see Och & Ney (2000)
Many researchers use the freely available GIZA++ tool to
automatically extract word alignments
Words

- Word-by-word translation
- Word alignments
  - Align words by hand
tedious and time-consuming
  - Automatic alignment
    uses EM on a parallel corpus;
see Och & Ney (2000)
Many researchers use the freely available GIZA++ tool to automatically extract word alignments
- Word alignments can be used in more sophisticated translation models.
Phrases

- Phrases may work better than words
Phrases

- Phrases may work better than words
- Phrase-based translation
Phrases

- Phrases may work better than words
- Phrase-based translation
  - Phrase table
Phrases

- Phrases may work better than words
- Phrase-based translation
  - Phrase table
  - Reordering model
Phrases

- Phrases may work better than words
- Phrase-based translation
  - Phrase table
  - Reordering model
- Phrases can be automatically extracted from word alignments.
Statistical Models

- Traditional noisy-channel approach
Statistical Models

- Traditional noisy-channel approach

\[
\arg\max_e P(e \mid f) = \arg\max_e P(e, f)
\]
Statistical Models

- Traditional noisy-channel approach

\[
\arg\max_e P(e | f) = \arg\max_e P(e, f)
\]

\[
\arg\max_e P(e | f) = \arg\max_e P(e) \times P(f | e)
\]
Statistical Models

- Traditional noisy-channel approach

\[
\arg\max_e P(e \mid f) = \arg\max_e P(e, f)
\]

\[
\arg\max_e P(e \mid f) = \arg\max_e P(e) \times P(f \mid e)
\]

- Log-linear approach
Statistical Models

- Traditional noisy-channel approach

\[
\arg \max_e P(e \mid f) = \arg \max_e P(e, f)
\]

\[
\arg \max_e P(e \mid f) = \arg \max_e P(e) \times P(f \mid e)
\]

- Log-linear approach

\[
weight = \prod_i \phi_i^{\lambda_i}
\]
So what features should our log-linear model use?
Log-linear features

So what features should our log-linear model use?

- $P_{lm}(e)$
- $P(f | e)$

Features used in noisy channel approach...
Log-linear features

So what features should our log-linear model use?

- $P_{lm}(e)$
- $P(f | e)$
- $P(e | f)$
- $P(f_w | e_w)$
- $P(e_w | f_w)$

Features used in noisy channel approach...
...and other features empirically found to be useful!
Hierarchical phrase-based translation

- Phrase-based translation
  - Phrase table
  - Reordering model
Hierarchical phrase-based translation

- Phrase-based translation
  - Phrase table
  - Reordering model
- What if we treat translation as a parsing task?
Hierarchical phrase-based translation

- Phrase-based translation
  - Phrase table
  - Reordering model
- What if we treat translation as a parsing task?
  → Phrase table becomes synchronous context free rules
Hierarchical phrase-based translation

- Phrase-based translation
  - Phrase table
  - Reordering model
- What if we treat translation as a parsing task?
  - Phrase table becomes synchronous context free rules
  - Reordering model becomes implicit in rule applications
Hierarchical phrase-based translation model
Hierarchical phrase-based translation model

- Database of synchronous context-free rules
Hierarchical phrase-based translation model

- Database of synchronous context-free rules
- Only two nonterminals!!!
Hierarchical phrase-based translation model

- Database of synchronous context-free rules
- Only two nonterminals!!!
  - X
Hierarchical phrase-based translation model

- Database of synchronous context-free rules
- Only two nonterminals!!!
  - X Used in extracted grammar rules
Hierarchical phrase-based translation model

- Database of synchronous context-free rules
- Only two nonterminals!!!
  - X Used in extracted grammar rules
  - S
Hierarchical phrase-based translation model

- Database of synchronous context-free rules
- Only two nonterminals!!!
  - X Used in extracted grammar rules
  - S Allows for serial combination of phrases
Our System

- Open source implementation of Chiang (2007)
Our System

- Open source implementation of Chiang (2007)
- Implemented in Java
Our System

- Open source implementation of Chiang (2007)
- Implemented in Java
- Designed to be easily extended
Our System

- Open source implementation of Chiang (2007)
- Implemented in Java
- Designed to be easily extended
- Data structures map onto the hypergraph architecture of Huang & Chiang (2005)
Our System

- Open source implementation of Chiang (2007)
- Implemented in Java
- Designed to be easily extended
- Data structures map onto the hypergraph architecture of Huang & Chiang (2005)
  - This allows n-best lists to be easily obtained
Our System

- Open source implementation of Chiang (2007)
- Implemented in Java
- Designed to be easily extended
- Data structures map onto the hypergraph architecture of Huang & Chiang (2005)
  - This allows n-best lists to be easily obtained
  - N-best lists are needed during parameter tuning
Our System

- Open source implementation of Chiang (2007)
- Implemented in Java
- Designed to be easily extended
- Data structures map onto the hypergraph architecture of Huang & Chiang (2005)
  - This allows n-best lists to be easily obtained
  - N-best lists are needed during parameter tuning
- Uses off-the-shelf minimum error rate trainer for log-linear parameter training
Our System

What make this system unique?
Our System

What make this system unique?

- There are two other known implementations of a hierarchical phrase-based system
Our System

What make this system unique?

▷ There are two other known implementations of a hierarchical phrase-based system
▷ Of the other two,...,
Our System

What make this system unique?

- There are two other known implementations of a hierarchical phrase-based system
- Of the other two,
  - CMU SAMT - Open source, but doesn't implement cube pruning algorithm. C++
Our System

What make this system unique?

- There are two other known implementations of a hierarchical phrase-based system
- Of the other two, ...
  - CMU SAMT - Open source, but doesn’t implement cube pruning algorithm. C++
  - Chiang’s Hiero - Closed source, does implement cube pruning algorithm. Python
Our System

What make this system unique?

- There are two other known implementations of a hierarchical phrase-based system
- Of the other two,...
  - CMU SAMT - Open source, but doesn’t implement cube pruning algorithm. C++
  - Chiang’s Hiero - Closed source, does implement cube pruning algorithm. Python
- Our system...
Our System

What make this system unique?

- There are two other known implementations of a hierarchical phrase-based system
- Of the other two,…,
  - CMU SAMT - Open source, but doesn’t implement cube pruning algorithm. C++
  - Chiang’s Hiero - Closed source, does implement cube pruning algorithm. Python

- Our system…
  - Open source
Our System

What make this system unique?

▶ There are two other known implementations of a hierarchical phrase-based system
▶ Of the other two…,
  ▶ CMU SAMT - Open source, but doesn’t implement cube pruning algorithm. C++
  ▶ Chiang’s Hiero - Closed source, does implement cube pruning algorithm. Python
▶ Our system…
  ▶ Open source
  ▶ Implements cube pruning algorithm.
Our System

What make this system unique?

- There are two other known implementations of a hierarchical phrase-based system
- Of the other two...
  - CMU SAMT - Open source, but doesn’t implement cube pruning algorithm. C++
  - Chiang’s Hiero - Closed source, does implement cube pruning algorithm. Python
- Our system...
  - Open source
  - Implements cube pruning algorithm.
  - Java
Our System

Code available through anonymous svn at http://sf.net/projects/nlp-parsers

Questions?

lane@cs.umn.edu